08/434,105

B31

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/82, 15/32, 15/60
C12N 15/54, 5/10, A01H 5/00
A01N 63/02

A1

(11) International Publication Number: WO 93/07278

(43) International Publication Date: 15 April 1993 (15.04.93)

(21) International Application Number: PCT/US92/08476

(22) International Filing Date: 5 October 1992 (05.10.92)

(30) Priority data:
772,027 4 October 1991 (04.10.91) US
951,715 25 September 1992 (25.09.92) US

(60) Parent Application or Grant
(63) Related by Continuation
US 951,715 (CON)
Filed on 25 September 1992 (25.09.92)

(71) Applicant *(for all designated States except US):* CIBA-GEIGY AG [CH/CH]; Klybeckstrasse 141, CH-4002 Basle (CH).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : KOZIEL, Michael, G. [US/US]; 509 Carolyn Court, Cary, NC 27511 (US). DESAI, Nalini, M. [US/US]; 107 Silverwood Lane, Cary, NC 27511 (US). LEWIS, Kelly, S. [US/US]; 1508 Rutoni Drive, Hillsborough, NC 27278 (US). KRAMER, Vance, C. [US/US]; 608 Dana Court, Hillsborough, NC 27278 (US). WARREN, Gregory, W. [US/US]; 324 Bond Lake Drive, Cary, NC 27513 (US). EVOLA, Stephen, V. [US/US]; 1013 E. Beechmont Circle, Apex, NC 27502 (US). CROSSLAND, Lyle, D. [US/US]; 108 Pinchot Lane, Chapel Hill, NC 27514 (US). WRIGHT, Martha, S. [US/US]; 106 Prince William Lane, Cary, NC 27511 (US). MERLIN, Ellis, J. [US/US]; 8605 Stonegate Drive, Raleigh, NC 27615 (US). LAUNIS, Karen, L. [US/US]; 104 Saddleview Lane, Franklinton, NC 27525 (US). ROTHSTEIN, Steven, J. [US/US]; 15 Fieldstone Road, Guelph, Ontario N1A 1A5 (US).

(74) Agent: FOLEY, Shawn, P.; Ciba-Geigy AG, P.O. Box 12257, Research Triangle Park, NC 27709 (US).

(81) Designated States: AU, BB, BG, BR, CA, CS, FI, HU, JP, KP, KR, LK, MG, MN, MW, NO, PL, RO, RU, SD, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SYNTHETIC DNA SEQUENCE HAVING ENHANCED INSECTICIDAL ACTIVITY IN MAIZE

(57) Abstract

DNA sequences optimized for expression in plants are disclosed. The DNA sequences preferably encode for an insecticidal polypeptides, particularly insecticidal proteins from *Bacillus thuringiensis*. Plant promoters, particular tissue-specific and tissue-preferred promoters are also provided. Additionally disclosed are transformation vectors comprising said DNA sequences. The transformation vectors demonstrate high levels of insecticidal activity when transformed into maize.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	LI	Liechtenstein	SK	Slovak Republic
CI	Côte d'Ivoire	LK	Sri Lanka	SN	Senegal
CM	Cameroon	LU	Luxembourg	SU	Soviet Union
CS	Czechoslovakia	MC	Monaco	TD	Chad
CZ	Czech Republic	MG	Madagascar	TG	Togo
DE	Germany	ML	Mali	UA	Ukraine
DK	Denmark	MN	Mongolia	US	United States of America
ES	Spain			VN	Viet Nam
FI	Finland				

SYNTHETIC DNA SEQUENCE HAVING ENHANCED INSECTICIDAL ACTIVITY IN MAIZE

This application is a continuation in part application of U.S. serial no. 772,027 filed October 4, 1991, which disclosure is herein incorporated in its entirety.

Field of the Invention

The present invention relates to DNA sequences encoding insecticidal proteins, and expression of these sequences in plants.

Background of the Invention

Expression of the insecticidal protein (IP) genes derived from *Bacillus thuringiensis* (Bt) in plants has proven extremely difficult. Attempts have been made to express chimeric promoter/Bt IP gene combinations in plants. Typically, only low levels of protein have been obtained in transgenic plants. *See*, for example, Vaeck et al., *Nature* 328:33-37, 1987; Barton et al., *Plant Physiol.* 85:1103-1109, 1987; Fischoff et al., *Bio/Technology* 5:807-813, 1987.

One postulated explanation for the cause of low expression is that fortuitious transcription processing sites produce aberrant forms of Bt IP mRNA transcript. These aberrantly processed transcripts are non-functional in a plant, in terms of producing an insecticidal protein. Possible processing sites include polyadenylation sites, intron splicing sites, transcriptional termination signals and transport

signals. Most genes do not contain sites that will deleteriously affect gene expression in that gene's normal host organism. However, the fortuitous occurrence of such processing sites in a coding region might complicate the expression of that gene in transgenic hosts. For example, the coding region for the Bt insecticidal crystal protein gene derived from Bacillus thuringiensis strain kurstaki (GENBANK BTHKURHD, accession M15271, B. thuringiensis var. kurstaki, HD-1; Geiser et al. Gene 48:109-118 (1986)) as derived directly from Bacillus thuringiensis, might contain sites which prevent this gene from being properly processed in plants.

Further difficulties exist when attempting to express Bacillus thuringiensis protein in an organism such as a plant. It has been discovered that the codon usage of a native Bt IP gene is significantly different from that which is typical of a plant gene. In particular, the codon usage of a native Bt IP gene is very different from that of a maize gene. As a result, the mRNA from this gene may not be efficiently utilized. Codon usage might influence the expression of genes at the level of translation or transcription or mRNA processing. To optimize an insecticidal gene for expression in plants, attempts have been made to alter the gene to resemble, as much as possible, genes naturally contained within the host plant to be transformed.

Adang et al., EP 0359472 (1990), relates to a synthetic Bacillus thuringiensis tenebrionis (Btt) gene which is 85% homologous to the native Btt gene and which is designed to have

an A+T content approximating that found in plants in general. Table 1 of Adang et al. show the codon sequence of a synthetic Btt gene which was made to resemble more closely the normal codon distribution of dicot genes. Adang et al. state that a synthetic gene coding for IP can be optimized for enhanced expression in monocot plants through similar methods, presenting the frequency of codon usage of highly expressed monocot proteins in Table 1. At page 9, Adang et al. state that the synthetic Btt gene is designed to have an A+T content of 55% (and, by implication, a G+C content of 45%). At page 20, Adang et al. disclose that the synthetic gene is designed by altering individual amino acid codons in the native Bt gene to reflect the overall distribution of codons preferred by dicot genes for each amino acid within the coding region of the gene. Adang et al. further state that only some of the native Btt gene codons will be replaced by the most preferred plant codon for each amino acid, such that the overall distribution of codons used in dicot proteins is preserved.

Fischhoff et al., EP 0 385 962 (1990), relates to plant genes encoding the crystal protein toxin of Bacillus thuringiensis. At table V, Fischhoff et al. disclose percent usages for codons for each amino acid. At page 8, Fischoff et al. suggest modifying a native Bt gene by removal of putative polyadenylation signals and ATTTA sequences. Fischoff et al. further suggest scanning the native Bt gene sequence for regions with greater than four consecutive adenine or thymine nucleotides to identify putative plant polyadenylation signals.

Fischoff et al. state that the nucleotide sequence should be altered if more than one putative polyadenylation signal is identified within ten nucleotides of each other. At page 9, Fischoff et al. state that efforts should be made to select codons to preferably adjust the G+C content to about 50%.

Perlak et al., PNAS USA, 88:3324-3328 (1991), relates to modified coding sequences of the Bacillus thuringiensis cryIA(b) gene, similar to those shown in Fischoff et al. As shown in table 1 at page 3325, the partially modified cryIA(b) gene of Perlak et al. is approximately 96% homologous to the native cryIA(b) gene (1681 of 1743 nucleotides), with a G+C content of 41%, number of plant polyadenylation signal sequences (PPSS) reduced from 18 to 7 and number of ATTTA sequences reduced from 13 to 7. The fully modified cryIA(b) gene of Perlak et al. is disclosed to be fully synthetic (page 3325, column 1). This gene is approximately 79% homologous to the native cryIA(b) gene (1455 of 1845 nucleotides), with a G+C content of 49%, number of plant polyadenylation signal sequences (PPSS) reduced to 1 and all ATTTA sequences removed.

Barton et al., EP 0431 829 (1991), relates to the expression of insecticidal toxins in plants. At column 10, Barton et al. describe the construction of a synthetic AaIT insect toxin gene encoding a scorpion toxin using the most preferred codon for each amino acid according to the chart shown in Figure 1 of the document.

Summary of the Invention

The present invention is drawn to methods for enhancing

expression of heterologous genes in plant cells. Generally, a gene or coding region of interest is constructed to provide a plant specific preferred codon sequence. In this manner, codon usage for a particular protein is altered to increase expression in a particular plant. Such plant optimized coding sequences can be operably linked to promoters capable of directing expression of the coding seqence in a plant cell.

Specifically, it is one of the objects of the present invention to provide synthetic insecticidal protein genes which have been optimized for expression in plants.

It is another object of the present invention to provide synthetic Bt insecticidal protein genes to maximize the expression of Bt proteins in a plant, preferably in a maize plant. It is one feature of the present invention that a synthetic Bt IP gene is constructed using the most preferred maize codons, except for alterations necessary to provide ligation sites for construction of the full synthetic gene.

According to the above objects, we have synthesized Bt insecticidal crystal protein genes in which the codon usage has been altered in order to increase expression in plants, particularly maize. However, rather than alter the codon usage to resemble a maize gene in terms of overall codon distribution, we have optimized the codon usage by using the codons which are most preferred in maize (maize preferred codons) in the synthesis of the synthetic gene. The optimized maize preferred codon usage is effective for expression of high

levels of the Bt insecticidal protein. This might be the result of maximizing the amount of Bt insecticidal protein translated from a given population of messenger RNAs. The synthesis of a Bt IP gene using maize preferred codons also tends to eliminate fortuitous processing sites that might occur in the native coding sequence. The expression of this synthetic gene is significantly higher in maize cells than that of the native IP Bt gene.

Preferred synthetic, maize optimized DNA sequences of the present invention derive from the protein encoded by the cryIA(b) gene in Bacillus thuringiensis var. kurstaki, HD-1; Geiser et al., Gene, 48:109-118 (1986) or the cryIB gene (AKA Crya4 gene) described by Brizzard and Whiteley, Nuc. Acids. Res., 16:2723 (1988). The DNA sequence of the native kurstaki HD-1 cryIA(b) gene is shown as Sequence 1. These proteins are active against various lepidopteran insects, including Ostrinia nubilalis, the European Corn Borer.

While the present invention has been exemplified by the synthesis of maize optimized Bt protein genes, it is recognized that the method can be utilized to optimize expression of any protein in plants.

The instant optimized genes can be fused with a variety of promoters, including constitutive, inducible, temporally regulated, developmentally regulated, tissue-preferred and tissue-specific promoters to prepare recombinant DNA molecules, i.e., chimeric genes. The maize optimized gene (coding sequence) provides substantially higher levels of expression in

a transformed plant, when compared with a non-maize optimized gene. Accordingly, plants resistant to Coleopteran or Lepidopteran pests, such as European corn borer and sugarcane borer, can be produced.

It is another object of the present invention to provide tissue-preferred and tissue-specific promoters which drive the expression of an operatively associated structural gene of interest in a specific part or parts of a plant to the substantial exclusion of other parts.

It is another object of the present invention to provide pith-preferred promoters. By "pith-preferred," it is intended that the promoter is capable of directing the expression of an operatively associated structural gene in greater abundance in the pith of a plant than in the roots, outer sheath, and brace roots, and with substantially no expression in seed.

It is yet another object of this invention to provide pollen-specific promoters. By "pollen-specific," it is intended that the promoter is capable of directing the expression of an operatively associated structural gene of interest substantially exclusively in the pollen of a plant, with negligible expression in any other plant part. By "negligible," it is meant functionally insignificant.

It is yet another object of the present invention to provide recombinant DNA molecules comprising a tissue-preferred promoter or tissue-specific promoter operably associated or linked to a structural gene of interest, particularly a

structural gene encoding an insecticidal protein, and expression of the recombinant molecule in a plant.

It is a further object of the present invention to provide transgenic plants which express at least one structural gene of interest operatively in a tissue-preferred or tissue-specific expression pattern.

In one specific embodiment of the invention disclosed and claimed herein, the tissue-preferred or tissue-specific promoter is operably linked to a structural gene encoding an insecticidal protein, and a plant is stably transformed with at least one such recombinant molecule. The resultant plant will be resistant to particular insects which feed on those parts of the plant in which the gene(s) is(are) expressed. Preferred structural genes encode B.t. insecticidal proteins. More preferred are maize optimized B.t. IP genes.

Brief Description of the Figures

Fig. 1 is a comparison of the full-length native Bt cryIA(b) gene [BTHKURHD], a full-length synthetic maize optimized Bt cryIA(b) gene [flsynbt.fin] and a truncated synthetic maize optimized Bt cryIA(b) gene [bssyn]. This figure shows that the full-length synthetic maize optimized cryIA(b) gene sequence matches that of the native cryIA(b) gene at about 2354 out of 3468 nucleotides (approximately 68% homology).

Fig. 2 is a comparison of the truncated native Bt cryIA(b) gene [BTHKURHD] and a truncated synthetic maize

optimized Bt gene [bssyn]. This figure shows that the truncated synthetic maize optimized cryIA(b) gene sequence matches that of the native cryIA(b) gene at about 1278 out of 1947 nucleotides (approximately 66% homology).

Fig. 3 is a comparison of the pure maize optimized Bt gene sequence [syn1T.mze] with a truncated synthetic maize optimized Bt gene [bssyn] and a full-length synthetic maize optimized Bt gene modified to include restriction sites for facilitating construction of the gene [synful.mod]. This figure shows that the truncated synthetic maize optimized cryIA(b) gene sequence matches that of the pure maize optimized cryIA(b) gene at 1913 out of 1947 nucleotides (approximately 98% homology).

Fig. 4 is a comparison of a native truncated Bt cryIA(b) gene [BTHKURHD] with a truncated synthetic cryIA(b) gene described in Perlak et al., PNAS USA, 88:3324-3328 (1991) [PMONBT] and a truncated synthetic maize optimized Bt gene [bssyn]. This figure shows that the PMONBT gene sequence matches that of the native cryIA(b) gene at about 1453 out of 1845 nucleotides (approximately 79% homology), while the truncated synthetic maize optimized Bt cryIA(b) gene matches the native cryIA(b) gene at about 1209 out of 1845 nucleotides (approximately 66% homology).

Fig. 5 is a comparison of a truncated synthetic cryIA(b) gene described in Perlak et al., PNAS USA, 88:3324-3328 (1991) [PMONBT] and a truncated synthetic maize optimized Bt cryIA(b) gene [bssyn]. This figure shows that the

PMONBT gene sequence matches that of the truncated synthetic maize optimized Bt cryIA(b) gene at about 1410 out of 1845 nucleotides (approximately 77% homology).

Fig. 6 is a full-length, maize optimized CryIB gene.

Fig. 7 is a full-length, hybrid, partially maize optimized DNA sequence of a CryIA(b) gene which is contained in pCIB4434. The synthetic region is from nucleotides 1-1938 (amino acids 1-646), and the native region is from nucleotides 1939-3468 (amino acids 647-1155). The fusion point between the synthetic and native coding sequences is indicated by a slash (/) in the sequence.

Fig. 8 is a map of pCIB4434.

Fig. 9 is a full-length, hybrid, maize optimized DNA sequence encoding a heat stable CryIA(b) protein, contained in pCIB5511.

Fig. 10 is a map of pCIB5511.

Fig. 11 is a full-length, hybrid, maize optimized DNA sequence encoding a heat stable CryIA(b) protein, contained in pCIB5512.

Fig. 12 is a map of pCIB5512.

Fig. 13 is a full-length, maize optimized DNA sequence encoding a heat stable CryIA(b) protein, contained in pCIB5513.

Fig. 14 is a map of pCIB5513.

Fig. 15 is a full-length, maize optimized DNA sequence encoding a heat-stable CryIA(b) gene, contained in pCIB5514.

Fig. 16 is a map of pCIB5514.

Fig. 17 is a map of pCIB4418.

Fig. 18 is a map of pCIB4420.

Fig. 19 is a map of pCIB4429.

Fig. 20 is a map of pCIB4431.

Fig. 21 is a map of pCIB4428.

Fig. 22 is a map of pCIB4430.

Fig. 23A is a table containing data of cryIA(b) protein levels in transgenic maize.

Fig. 23B is a table which summarizes results of bioassays of Ostrinia and Diatraea on leaf material from maize progeny containing a maize optimized CryIA(b) gene.

Fig. 23C is a table containing data of cryIA(b) protein levels in transgenic maize.

Fig. 23D is a table which summarizes the results of bioassays of Ostrinia and Diatraea on leaf material from maize progeny containing a synthetic Bt. maize gene operably linked to a pith promoter.

Fig. 23E is a table containing data on expression of the cryIA(b) gene in transgenic maize using the pith-preferred promoter.

Fig. 24 is a complete genomic DNA sequence of a maize tryptophan synthase-alpha subunit gene. Introns, exons, transcription and translation starts, start and stop of cDNA are shown. $ = start and end of cDNA; +1 = transcription start; 73******* = primer extension primer; +1 = start of translation; +++ = stop codon; bp 1495-99 = CCAAT Box; bp 1593-1598 = TATAA Box; bp 3720-3725 = poly A addition site; # above underlined sequences are PCR primers.

Figs. 25A, 25B, 25C and 25D are Northern blot analyses which show differential expression of the maize TrpA subunit gene in maize tissue at 2 hour, 4 hour, 18 hour, and 48 hour intervals, respectively, at -80•C with DuPont Cronex intensifying screens. P=pith; C=cob; BR=brace roots; ES=ear shank; LP=lower pith; MP=middle pith; UP=upper pith; S=seed; L=leaf; R=root; and P(upper left)=total pith.

Fig. 26 is a Northern blot analysis, the two left lanes of which show the maize TrpA gene expression in the leaf (L) and pith (P) of Funk inbred lines 211D and 5N984. The five right lanes indicate the absence of expression in Funk 211D seed total RNA. S(1, 2,3, 4 and 5)= seed at 1, 2, 3, 4 and 5 weeks post pollenation. L=leaf; P=pith; S#=seed # weeks post pollenation.

Fig. 27 is a Southern blot analysis of genomic DNA Funk line 211D, probed with maize TrpA cDNA 8-2 (pCIB5600), wherein B denotes BamHI, E denotes EcoRI, EV denotes EcoRV, H denotes HINDIII, and S denotes SacI. 1X, 5X and 10X denote reconstructed gene copy equivalents.

Fig. 28A is a primer extension analysis which shows the transcriptional start of the maize TrpA subunit gene and sequencing ladder. Lane +1 and +2 are 1X + 0.5X samples of primer extension reaction.

Fig. 28B is an analysis of RNase protection from +2 bp to +387 bp at annealing temperatures of 42•C, 48•C and 54•C, at a 16 hour exposure against film at -80•C with DuPont Cronex intensifying screens.

Fig. 29 is A map of the original Type II pollen-specific cDNA clone. The subcloning of the three EcoRI fragments into pBluescript vectors to create pCIB3168, pCIB3169 and II-.6 is illustrated.

Fig. 30 shows the DNA sequence of the maize pollen-specific calcium dependent protein kinase gene cDNA, as contained in the 1.0 kb and 0.5 kb fragments of the original Type II cDNA clone. The EcoRI site that divides the 1.0 kb and 0.5 kb fragments is indicated. This cDNA is not full length, as the mRNA start site maps 490 bp upstream of the end of the cDNA clone.

Fig. 31 illustrates the tissue-specific expression of the pollen CDPK mRNA. RNA from the indicated maize 211D tissues was denatured, electrophoresed on an agarose gel, transferred to nitrocellulose, and probed with the pollen CDPK cDNA 0.5 kb fragment. The mRNA is detectable only in the pollen, where a strong signal is seen.

Fig. 32 is an amino acid sequence comparison of the pollen CDPK derived protein sequence and the rat protein kinase 2 protein sequence disclosed in Tobimatsu et al., J. Biol. Chem. 263:16082-16086 (1988). The Align program of the DNAstar software package was used to evaluate the sequences. The homology to protein kinases occurs in the 5′ two thirds of the gene, i.e. in the 1.0 kb fragment.

Fig. 33 is an amino acid sequence comparison of the pollen CDPK derived protein sequence and the human calmodulin protein sequence disclosed in Fischer et al., J. Biol. Chem.

263:17055-17062 (1988). The homology to calmodulin occurs in the 3′ one third of the gene, i.e. in the 0.5 kb fragment.

Fig. 34 is an amino acid sequence comparison of the pollen CDPK derived protein sequence and soybean CDPK. The homology occurs over the entire gene.

Fig. 35 illustrates the sequence of the maize pollen-specific CDPK gene. 1.4 kb of sequence prior to the mRNA start site is shown. The positions of the seven exons and six introns are depicted under the corresponding DNA sequence. The site of polyadenylation in the cDNA clone is indicated.

Fig. 36 is a map of pCIB4433.

Fig. 37 is a full-length, hybrid, maize-optimized DNA sequence encoding a heat stable cryIA(b) protein.

Fig. 38 is a map of pCIB5515.

Description of the Sequences:

Sequence 1 is the DNA sequence of a full-length native Bt cryIA(b) gene.

Sequence 2 is the DNA sequence of a full-length pure maize optimized synthetic Bt cryIA(b) gene.

Sequence 3 is the DNA sequence of an approximately 2 Kb truncated synthetic maize optimized Bt cryIA(b) gene.

Sequence 4 is the DNA sequence of a full-length synthetic maize optimized Bt cryIA(b) gene.

Sequence 5 is the DNA sequence of an approximately 2 Kb synthetic Bt gene according to Perlak et al.

Detailed Description of the Invention

The following definitions are provided in order to

provide clarity with respect to the terms as they are used in the specification and claims to describe the present invention.

Maize preferred codon: Preferred codon refers to the preference exhibited by a specific host cell in the usage of nucleotide codons to specify a given amino acid. The preferred codon for an amino acid for a particular host is the single codon which most frequently encodes that amino acid in that host. The maize preferred codon for a particular amino acid may be derived from known gene sequences from maize. For example, maize codon usage for 28 genes from maize plants are listed in Table 4 of Murray et al., Nucleic Acids Research, 17:477-498 (1989), the disclosure of which is incorporated herein by reference. For instance, the maize preferred codon for alanine is GCC, since, according to pooled sequences of 26 maize genes in Murray et al., supra, that codon encodes alanine 36% of the time, compared to GCG (24%), GCA (13%), and GCT (27%).

Pure maize optimized sequence: An optimized gene or DNA sequence refers to a gene in which the nucleotide sequence of a native gene has been modified in order to utilize preferred codons for maize. For example, a synthetic maize optimized Bt cryIA(b) gene is one wherein the nucleotide sequence of the native Bt cryIA(b) gene has been modified such that the codons used are the maize preferred codons, as described above. A pure maize optimized gene is one in which the nucleotide sequence comprises 100 percent of the maize preferred codon sequences for a particular polypeptide. For example, the pure

maize optimized Bt cryIA(b) gene is one in which the nucleotide sequence comprises 100 percent maize preferred codon sequences and encodes a polypeptide with the same amino acid sequence as that produced by the native Bt cryIA(b) gene. The pure nucleotide sequence of the optimized gene may be varied to permit manipulation of the gene, such as by altering a nucleotide to create or eliminate restriction sites. The pure nucleotide sequence of the optimized gene may also be varied to eliminate potentially deleterious processing sites, such as potential polyadenylation sites or intron recognition sites.

It is recognized that "partially maize optimized," sequences may also be utilized. By partially maize optimized, it is meant that the coding region of the gene is a chimeric (hybrid), being comprised of sequences derived from a native insecticidal gene and sequences which have been optimized for expression in maize. A partially optimized gene expresses the insecticidal protein at a level sufficient to control insect pests, and such expression is at a higher level than achieved using native sequences only. Partially maize optimized sequences include those which contain at least about 5% optimized sequences.

<u>Full-length Bt Genes:</u> Refers to DNA sequences comprising the full nucleotide sequence necessary to encode the polypeptide produced by a native Bt gene. For example, the native Bt cryIA(b) gene is approximately 3.5 Kb in length and encodes a polypeptide which is approximately 1150 amino acids in length. A full-length synthetic cryIA(b) Bt gene would be

at least approximately 3.5 Kb in length.

Truncated Bt Genes: Refers to DNA sequences comprising less than the full nucleotide sequence necessary to encode the polypeptide produced by a native Bt gene, but which encodes the active toxin portion of the polypeptide. For example, a truncated synthetic Bt gene of approximately 1.9 Kb encodes the active toxin portion of the polypeptide such that the protein product exhibits insecticidal activity.

Tissue-preferred promoter: The term "tissue-preferred promoter" is used to indicate that a given regulatory DNA sequence will promote a higher level of transcription of an associated structural gene or DNA coding sequence, or of expression of the product of the associated gene as indicated by any conventional RNA or protein assay, or that a given DNA sequence will demonstrate some differential effect; i.e., that the transcription of the associated DNA sequences or the expression of a gene product is greater in some tissue than in all other tissues of the plant.

"Tissue-specific promoter" is used to indicate that a given regulatory DNA sequence will promote transcription of an associated coding DNA sequence essentially entirely in one or more tissues of a plant, or in one type of tissue, e.g. green tissue, while essentially no transcription of that associated coding DNA sequence will occur in all other tissues or types of tissues of the plant.

The present invention provides DNA sequences optimized for expression in plants, especially in maize plants. In a

preferred embodiment of the present invention, the DNA sequences encode the production of an insecticidal toxin, preferably a polypeptide sharing substantially the amino acid sequence of an insecticidal crystal protein toxin normally produced by Bacillus thuringiensis. The synthetic gene may encode a truncated or full-length insecticidal protein. Especially preferred are synthetic DNA sequences which encode a polypeptide effective against insects of the order Lepidoptera and Coleoptera, and synthetic DNA sequences which encode a polypeptide having an amino acid sequence essentially the same as one of the crystal protein toxins of Bacillus thuringiensis variety kurstaki, HD-1.

The present invention provides synthetic DNA sequences effective to yield high expression of active insecticidal proteins in plants, preferably maize protoplasts, plant cells and plants. The synthetic DNA sequences of the present invention have been modified to resemble a maize gene in terms of codon usage and G+C content. As a result of these modifications, the synthetic DNA sequences of the present invention do not contain the potential processing sites which are present in the native gene. The resulting synthetic DNA sequences (synthetic Bt IP coding sequences) and plant transformation vectors containing this synthetic DNA sequence (synthetic Bt IP genes) result in surprisingly increased expression of the synthetic Bt IP gene, compared to the native Bt IP gene, in terms of insecticidal protein production in plants, particularly maize. The high level of expression

results in maize cells and plants that exhibit resistance to lepidopteran insects, preferably European Corn Borer and Diatrea saccharalis, the Sugarcane Borer.

The synthetic DNA sequences of the present invention are designed to encode insecticidal proteins from Bacillus thuringiensis, but are optimized for expression in maize in terms of G+C content and codon usage. For example, the maize codon usage table described in Murray et al., supra, is used to reverse translate the amino acid sequence of the toxin produced by the Bacillus thuringiensis subsp. kurstaki HD-1 cryIA(b) gene, using only the most preferred maize codons. The reverse translated DNA sequence is referred to as the pure maize optimized sequence and is shown as Sequence 4. This sequence is subsequently modified to eliminate unwanted restriction endonuclease sites, and to create desired restriction endonuclease sites. These modifications are designed to facilitate cloning of the gene without appreciably altering the codon usage or the maize optimized sequence. During the cloning procedure, in order to facilitate cloning of the gene, other modifications are made in a region that appears especially susceptible to errors induced during cloning by the polymerase chain reaction (PCR). The final sequence of the maize optimized synthetic Bt IP gene is shown in Sequence 2. A comparision of the maize optimized synthetic Bt IP gene with the native kurstaki cryIA(b) Bt gene is shown in Fig. 1.

In a preferred embodiment of the present invention, the protein produced by the synthetic DNA sequence is effective

against insects of the order Lepidoptera or Coleoptera. In a more preferred embodiment, the polypeptide encoded by the synthetic DNA sequence consists essentially of the full-length or a truncated amino acid sequence of an insecticidal protein normally produced by Bacillus thuringiensis var. kurstaki HD-1. In a particular embodiment, the synthetic DNA sequence encodes a polypeptide consisting essentially of a truncated amino acid sequence of the Bt CryIA(b) protein.

The insecticidal proteins of the invention are expressed in a plant in an amount sufficient to control insect pests, i.e. insect controlling amounts. It is recognized that the amount of expression of insecticidal protein in a plant necessary to control insects may vary depending upon species of plant, type of insect, environmental factors and the like. Generally, the insect population will be kept below the economic threshold which varies from plant to plant. For example, to control European corn borer in maize, the economic threshold is .5 eggmass/plant which translates to about 10 larvae/plant.

The methods of the invention are useful for controlling a wide variety of insects including but not limited to rootworms, cutworms, armyworms, particularly fall and beet armyworms, wireworms, aphids, corn borers, particularly European corn borers, sugarcane borer, lesser corn stalk borer, Southwestern corn borer, etc.

In a preferred embodiment of the present invention, the synthetic coding DNA sequence optimized for expression in

maize comprises a G+C percentage greater than that of the native cryIA(b) gene. It is preferred that the G+C percentage be at least about 50 percent, and more preferably at least about 60 percent. It is especially preferred that the G+C percent be about 64 percent.

In another preferred embodiment of the present invention, the synthetic coding DNA sequence optimized for expression in maize comprises a nucleotide sequence having at least about 90 percent homology with the "pure" maize optimized nucleotide sequence of the native Bacillus thuringiensis cryIA(b) protein, more preferably at least about 95 percent homology, and most preferably at least about 98 percent.

Other preferred embodiments of the present invention include synthetic DNA sequences having essentially the DNA sequence of Sequence ID No. 4, as well as mutants or variants thereof; transformation vectors comprising essentially the DNA sequence of Sequence ID No. 4; and isolated DNA sequences derived from the plasmids pCIB4406, pCIB4407, pCIB4413, pCIB4414, pCIB4416, pCIB4417, pCIB4418, pCIB4419, pCIB4420, pCIB4421, pCIB4423, pCIB4434, pCIB4429, pCIB4431, pCIB4433. Most preferred are isolated DNA sequences derived from the plasmids pCIB4418 and pCIB4420, pCIB4434, pCIB4429, pCIB4431, and pCIB4433.

In order to construct one of the maize optimized DNA sequences of the present invention, synthetic DNA oligonucleotides are made with an average length of about 80 nucleotides. These oligonucleotides are designed to hybridize

to produce fragments comprising the various quarters of the truncated toxin gene. The oligonucleotides for a given quarter are hybridized and amplified using PCR. The quarters are then cloned and the cloned quarters are sequenced to find those containing the desired sequences. In one instance, the fourth quarter, the hybridized oligonucleotides are cloned directly without PCR amplification. Once all clones of four quarters are found which contain open reading frames, an intact gene encoding the active insecticidal protein is assembled. The assembled gene may then be tested for insecticidal activity against any insect of interest including the European Corn Borer (ECB) and the sugarcane borer. (Examples 5A and 5B, respectively). When a fully functional gene is obtained, it is again sequenced to confirm its primary structure. The fully functional gene is found to give 100% mortality when bioassayed against ECB. The fully functional gene is also modified for expression in maize.

The maize optimized gene is tested in a transient expression assay, e.g. a maize transient expression assay. The native Bt cryIA(b) coding sequence for the active insecticidal toxin is not expressed at a detectable level in a maize transient expression system. Thus, the level of expression of the synthesized gene can be determined. By the present methods, expression of a protein in a transformed plant can be increased at least about 100 fold to about 50,000 fold, more specifically at least about 1,000 fold to at least about 20,000 fold.

Increasing expression of an insecticial gene to an effective level does not require manipulation of a native gene along the entire sequence. Effective expression can be achieved by manipulating only a portion of the sequences necessary to obtain increased expression. A full-length, maize optimized CryIA(b) gene may be prepared which contains a protein of the native CryIA(b) sequence. For example, Figure 7 illustrates a full-length, maize optimized CryIA(b) gene which is a synthetic-native hybrid. That is, about 2kb of the gene (nucleotides 1-1938) is maize optimized, i.e. synthetic. The remainder, C-terminal nucleotides 647-1155, are identical to the corresponding sequence native of the CryIA(b) gene. Construction of the illustrated gene is described in Example 6, below.

It is recognized that by using the methods described herein, a variety of synthetic/native hybrids may be constructed and tested for expression. The important aspect of hybrid construction is that the protein is produced in sufficient amounts to control insect pests. In this manner, critical regions of the gene may be identified and such regions synthesized using preferred codons. The synthetic sequences can be linked with native sequences as demonstrated in the Examples below. Generally, N-terminal portions or processing sites can be synthesized and substituted in the native coding sequence for enhanced expression in plants.

In another embodiment of the present invention, the maize optimized genes encoding cryIA(b) protein may be

manipulated to render the encoded protein more heat stable or temperature stable compared to the native cryIA(b) protein. It has been shown that the cryIA(b) gene found in Bacillus thuringiensis kurstaki HD-1 contains a 26 amino acid deletion, when compared with the cryIA(a) and cryIA(c) proteins, in the -COOH half of the protein. This deletion leads to a temperature-sensitive cryIA(b) protein. See M. Geiser, EP 0 440 581, entitled "Temperaturstabiles Bacillus thuringiensis-Toxin". Repair of this deletion with the corresponding region from the cryIA(a) or cryIA(c) protein improves the temperature stability of the repaired protein. Constructs of the full-length modified cryIA(b) synthetic gene are designed to insert sequences coding for the missing amino acids at the appropriate place in the sequence without altering the reading frame and without changing the rest of the protein sequence. The full-length synthetic version of the gene is assembled by synthesizing a series of double-stranded DNA cassettes, each approximately 300 bp in size, using standard techniques of DNA synthesis and enzymatic reactions. The repaired gene is said to encode a "heat stable" or "temperature-stable" cryIA(b) protein, since it retains more biological activity than its native counterpart when exposed to high temperatures. Specific sequences of maize optimized, heat stable cryIA(b) genes encoding temperature stable proteins are set forth in Figs. 9, 11, 13, and 15, and are also described in Example 7, below.

The present invention encompasses maize optimized

coding sequences encoding other polypeptides, including those of other Bacillus thuringiensis insecticidal polypeptides or insecticidal proteins from other sources. For example, cryIB genes can be maize optimized, and then stably introduced into plants, particularly maize. The sequence of a maize optimized cryIB gene constructed in accordance with the present invention is set forth in Fig. 6.

Optimizing a Bt IP gene for expression in maize using the maize preferred codon usage according to the present invention results in a significant increase in the expression of the insecticidal gene. It is anticipated that other genes can be synthesized using plant codon preferences to improve their expression in maize or other plants. Use of maize codon preference is a likely method of optimizing and maximizing expression of foreign genes in maize. Such genes include genes used as selectable or scoreable markers in maize transformation, genes which confer herbicide resistance, genes which confer disease resistance, and other genes which confer insect resistance.

The synthetic cryIA(b) gene is also inserted into Agrobacterium vectors which are useful for transformation of a large variety of dicotyledenous plant species. (Example 44). Plants stably transformed with the synthetic cryIA(b) Agrobacterium vectors exhibit insecticidal activity.

The native Bt cryIA(b) gene is quite A+T rich. The G+C content of the full-length native Bt cryIA(b) gene is approximately 39%. The G+C content of a truncated native Bt

cryIA(b) gene of about 2 Kb in length is approximately 37%. In general, maize coding regions tend to be predominantly G+C rich. The modifications made to the Bt cryIA(b) gene result in a synthetic IP coding region which has greater than 50% G+C content, and has about 65% homology at the DNA level with the native cryIA(b) gene. The protein encoded by this synthetic CryIA(b) gene is 100% homologous with the native protein, and thus retains full function in terms of insect activity. The truncated synthetic CryIA(b) IP gene is about 2 Kb in length and the gene encodes the active toxin region of the native Bt kurstaki CryIA(b) insecticidal protein. The length of the protein encoded by the truncated synthetic CryIA(b) gene is 648 amino acids.

The synthetic genes of the present invention are useful for enhanced expression in transgenic plants, most preferably in transformed maize. The transgenic plants of the present invention may be used to express the insecticidal CryIA(b) protein at a high level, resulting in resistance to insect pests, preferably coleopteran or lepidopteran insects, and most preferably European Corn Borer (ECB) and Sugarcane Borer.

In the present invention, the DNA coding sequence of the synthetic maize optimized gene may be under the control of regulatory elements such as promoters which direct expression of the coding sequence. Such regulatory elements, for example, include monocot or maize and other monocot functional promoters to provide expression of the gene in various parts of the maize plant. The regulatory element may be constitutive. That is,

it may promote continuous and stable expression of the gene. Such promoters include but are not limited to the CaMV 35S promoter; the CaMV 19S promoter; A. tumefaciens promoters such as octopine synthase promoters, mannopine synthase promoters, nopaline synthase promoters, or other opine synthase promoters; ubiquitin promoters, actin promoters, histone promoters and tubulin promoters. The regulatory element may be a tissue-preferential promoter, that is, it may promote higher expression in some tissues of a plant than in others. Preferably, the tissue-preferential promoter may direct higher expression of the synthetic gene in leaves, stems, roots and/or pollen than in seed. The regulatory element may also be inducible, such as by heat stress, water stress, insect feeding or chemical induction, or may be developmentally regulated. Numerous promoters whose expression are known to vary in a tissue specific manner are known in the art. One such example is the maize phosphoenol pyruvate carboxylase (PEPC), which is green tissue-specific. See, for example, Hudspeth, R.L. and Grula, J.W., Plant Molecular Biology 12:579-589, 1989). Other green tissue-specific promoters include chlorophyll a/b binding protein promoters and RubisCO small subunit promoters.

The present invention also provides isolated and purified pith-preferred promoters. Preferred pith-preferred promoters are isolated from graminaceous monocots such as sugarcane, rice, wheat, sorghum, barley, rye and maize; more preferred are those isolated from maize plants.

In a preferred embodiment, the pith-preferred promoter

is isolated from a plant TrpA gene; in a most preferred embodiment, it is isolated from a maize TrpA gene. That is, the promoter in its native state is operatively associated with a maize tryptophan synthase-alpha subunit gene (hereinafter "TrpA"). The encoded protein has a molecular mass of about 38kD. Together with another alpha subnit and two beta subunits, TrpA forms a multimeric enzyme, tryptophan synthase. Each subunit can operate separately, but they function more efficiently together. TrpA catalyzes the conversion of indole glycerol phosphate to indole. Neither the maize TrpA gene nor the encoded protein had been isolated from any plant before Applicants' invention. The *Arabidopsis thaliana* tryptophan synthase beta subunit gene has been cloned as described Wright et al., *The Plant Cell*, 4:711-719 (1992). The instant maize TrpA gene has no homology to the beta subunit encoding gene.

The present invention also provides purified pollen-specific promoters obtainable from a plant calcium-dependent phosphate kinase (CDPK) gene. That is, in its native state, the promoter is operably linked to a plant CDPK gene. In a preferred embodiment, the promoter is isolated from a maize CDPK gene. By "pollen-specific," it is meant that the expression of an operatively associated structural gene of interest is substantially exclusively (i.e. essentially entirely) in the pollen of a plant, and is negligible in all other plant parts. By "CDPK," it is meant a plant protein kinase which has a high affinity for calcium, but not calmodulin, and requires calcium, but not calmodulin, for its

catalytic activity.

To obtain tissue-preferred or tissue specific promoters, genes encoding tissue specific messenger RNA (mRNA) can be obtained by differential screening of a cDNA library. For example, a pith-preferred cDNA can be obtained by subjecting a pith cDNA library to differential screening using cDNA probes obtained from pith and seed mRNA. See, Molecular Cloning, A Laboratory Manual, Sambrook et al. eds. Cold Spring Harbor Press: New York (1989).

Alternately, tissue specific promoters may be obtained by obtaining tissue specific proteins, sequencing the N-terminus, synthesizing oligonucleotide probes and using the probes to screen a cDNA library. Such procedures are exemplified in the Experimental section for the isolation of a pollen specific promoter.

The scope of the present invention in regard to the pith-preferred and pollen-specific promoters encompasses functionally active fragments of a full-length promoter that also are able to direct pith-preferred or pollen-specific transcription, respectively, of associated structural genes. Functionally active fragments of a promoter DNA sequence may be derived from a promoter DNA sequence, by several art-recognized procedures, such as, for example, by cleaving the promoter DNA sequence using restriction enzymes, synthesizing in accordance with the sequence of the promoter DNA sequence, or may be obtained through the use of PCR technology. See, e.g. Mullis et al., Meth. Enzymol. 155:335-350 (1987); Erlich (ed.), PCR

Technology, Stockton Press (New York 1989).

Further included within the scope of the instant invention are pith-preferred and pollen-specific promoters "equivalent" to the full-length promoters. That is, different nucleotides, or groups of nucleotides may be modified, added or deleted in a manner that does not abolish promoter activity in accordance with known procedures.

A pith-preferred promoter obtained from a maize TrpA gene is shown in Fig. 24. Those skilled in the art, with this sequence information in hand, will recognize that pith-preferred promoters included within the scope of the present invention can be obtained from other plants by probing pith libraries from these plants with probes derived from the maize TrpA structural gene. Probes designed from sequences that are highly conserved among TrpA subunit genes of various species, as discussed generally in Example 17, are preferred. Other pollen-specific promoters, which in their native state are linked to plant CDPK genes other than maize, can be isolated in similar fashion using probes derived from the conserved regions of the maize CDPK gene to probe pollen libraries.

In another embodiment of the present invention, the pith-preferred or pollen-specific promoter is operably linked to a DNA sequence, i.e. structural gene, encoding a protein of interest, to form a recombinant DNA molecule or chimeric gene. The phrase "operably linked to" has an art-recognized meaning; it may be used interchangeably with "operatively associated

with,""linked to," or "fused to".

The structural gene may be homologous or heterologous with respect to origin of the promoter and/or a target plant into which it is transformed. Regardless of relative origin, the associated DNA sequence will be expressed in the transformed plant in accordance with the expression properties of the promoter to which it is linked. Thus, the choice of associated DNA sequence should flow from a desire to have the sequence expressed in this fashion. Examples of heterologous DNA sequences include those which encode insecticidal proteins, e.g. proteins or polypeptides toxic or inhibitory to insects or other plant parasitic arthropods, or plant pathogens such as fungi, bacteria and nematodes. These heterologous DNA sequences encode proteins such as magainins, Zasloff, PNAS USA, 84:5449-5453 (1987); cecropins, Hultmark et al., Eur. J. Biochem. 127:207-217 (1982); attacins, Hultmark et al., EMBO J. 2:571-576 (1983); melittin, gramicidin S, Katsu et al., Biochem. Biophys. Acta, 939:57-63 (1988); sodium channel proteins and synthetic fragments, Oiki et al. PNAS USA, 85:2395-2397 (1988); the alpha toxin of Staphylococcus aureusm Tobkes et al., Biochem., 24:1915-1920 (1985); apolipoproteins and fragments thereof, Knott et al., Science 230:37 (1985); Nakagawa et al., J. Am. Chem. Soc., 107:7087 (1985); alamethicin and a variety of synthetic amphipathic peptides, Kaiser et al., Ann. Rev. Biophys. Biophys. Chem. 16:561-581 (1987); lectins, Lis et al., Ann. Rev. Biochem., 55:35-68 (1986); protease and amylase inhibitors; and insecticidal

ERROR: ioerror
OFFENDING COMMAND: image

STACK:

-dictionary-
-savelevel-